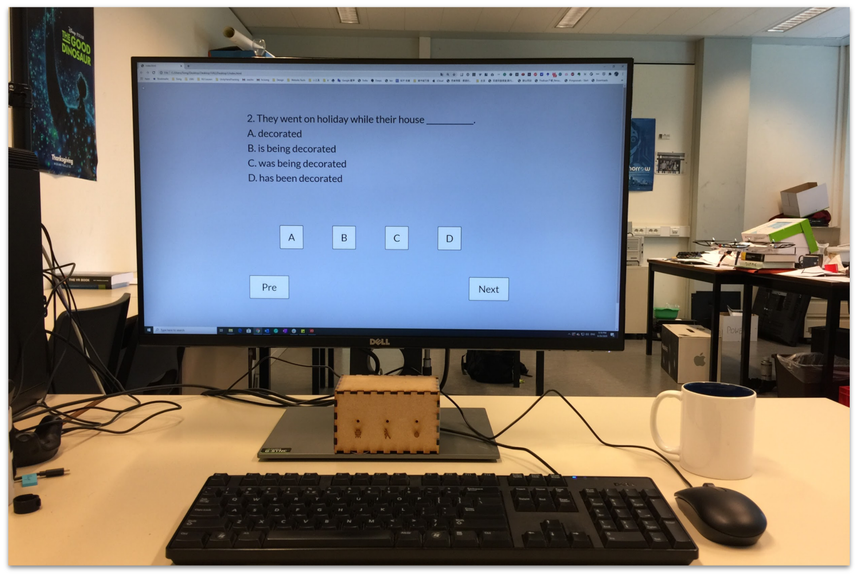
Using Heuristic Evaluation in Immersive Virtual Reality Evaluation

Xuesong Zhang
KU Leuven, Belgium

Adalberto L. Simeone
KU Leuven, Belgium



Participants view in RE and VE

BACKGROUND

Using Virtual Reality (VR) as a proxy to perform usability testing has already gained researchers attention.

RQ

Can conventional usability evaluation methods can directly be applied to VR evaluations?
Will they lead to similar insights when compared to equivalent real-world lab studies?

USER STUDY

We conducted a user study with nine participants to test the performance of evaluating a prototype of an imagined future technology in the real life and an immersive virtual environment by using a *Heuristic evaluation (HE).* This user study follows a between-subjects design. For each group, the independent variable is ENVIRONMENT: *Real environment (RE), Virtual environment (VE).*

HEALTH BOX

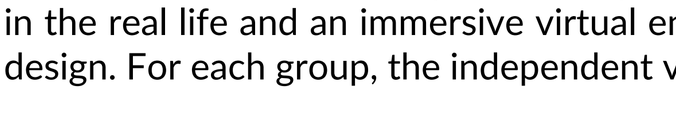
The Health Box is a smart device that detects the correctness of the user's posture, hydration level and exercise frequency and gives feedback to the user through three light indications. It is placed under the monitor. During the experiment, users acknowledge the exercise by tapping the health box. Likewise for the hydration levels, users touch the mug to indicate the drinking action.

TASK

Participants sit in front of a (virtual) desktop. They need to complete two English Tests during the experiment. In the process of answering one of the English tests, they need to respond the health box in the meantime. The order of the two English tests and with or without responding the health box was counterbalanced. See participant view in RE and VE above.

USER STUDY

Usability problems identified



VR unique:
- The light is not bright enough *(Implementation);*
- Icons are too small *(Implementation);*
- Blur icons *(Resolution);*
- The health box is out of view field *(FoV);*
- The health box should be placed somewhere else *(FoV);*

RE unique:
- The light is distracting *(FoV).*

Custom Questionnaire - Use Experience

(1: strongly disagree; 5: strongly agree)

"I enjoyed the experience."
VE (*MEAN* = 3.58) > RE (*MEAN* = 3.11), $p = 0.304$

"It was easy to pay attention to the work."
VE (*MEAN*= 3.47) > RE (*MEAN*= 2.77), $p = 0.129$

"It was annoying being notified while working."
VE (*MEAN* = 2.78) < RE (*MEAN* = 3.22), $p = 0.385$

"The posture detection of the device is very sensitive. "
VE (*MEAN* = 3.31) > RE (*MEAN* = 3.22), $p = 0.662$

Performance

We calculated the difference in the number of incorrect answers for each participant on the two English tests. We compared this difference between using the health box in VE (*MEAN* = 0.17) and in RE (*MEAN* = 1.37) with a Kruskal-Wallis H test. No statistically difference ($p = 0.276$) was found.

User Behaviour Mode

We observed that participants interacted in VR in a similar way to how they interacted in RE.

DISCUSSION AND CONCLUSION

Using the heuristic evaluation, the most usability problems present on a virtual equivalent are also identified on the physical prototype. We noticed that the implementation of the virtual counterparts did affect the experts' evaluation results. The low resolution and limited field of view of VR headsets may lead inspectors to report false positives. Thus we suggest that future work should focus on the impact of the representation fidelity in immersive virtual evaluations, in terms of the graphical visual aspect.

Contact: xuesong.zhang@kuleuven.be